Marie Thomas
Professor of Psychology
California State University San Marcos

**TBL Module: Categorical Data and Chi-Square**

**Table of contents**

* **Introduction (p. 1)**
* **Reading assignment and learning outcomes (p. 1)**
* **Readiness assurance test (pp. 3-4)**
* Plan of class (p. 2)
* **Introductory application activity (pp. 5-6)**
* Ending application activity (p. 7-8)

**Introduction**

This module is one I use in my advanced psychological statistics course. The name of the course is deceptive; it is actually intermediate level. I discovered early on that students often remembered little from their introductory course so my class has evolved to the point where I spend the first half of the semester reviewing introductory statistics topics at a deeper level; I then move to new material after the midterm.

Most of the topics in this module are covered in introductory statistics classes and the concepts are quite elementary. Yet I find that students have a great deal of trouble understanding these statistical techniques. It may be because it’s often the last topic covered in the introductory course and instructors may not have enough time to do it justice. Whatever the reason, this module requires more repetition of concepts than would seem necessary or appropriate for a class at this level. But I like the application activities, as do the students who find them very useful. This module typically requires 4 to 5 hours of class time.

**Reading assignment and learning outcomes**

I use a textbook for this course and, in the first half of the semester, cover approximately one chapter per week. The reading assignment for this module is Ch. 6 in David Howell’s (2013) *Statistical methods for psychology* (8th Ed.). Belmont, CA: Thomson.

I create a “reading guide” for each module which is, essentially, the learning outcomes. I don’t expect my students to understand the entire chapter without some help. I highlight the outcomes that will be covered by RAT questions, but students understand that for the midterm/final they are responsible for understanding all of the listed concepts.

* **I can describe the characteristics of the chi-square distribution.**
* **I can define the terms expected and observed frequencies.**
* **I can perform a chi-square goodness-of-fit test by hand and using statistical software, and can interpret the results.**
* **I can perform chi-square contingency table analysis (test of independence) by hand** and using statistical software**, and can interpret the results.**
* I can describe the problems associated with small expected frequencies.
* **I can describe the assumptions associated with chi-square.**
* I can describe one way to handle dependent or repeated measurements using chi-square.
* I can explain and demonstrate how to calculate d-family and r-family measures of effect size.
* I can explain and demonstrate how to calculate kappa.

**Plan of class**

* iRAT and tRAT
* Introductory application activity: What kind of chi-square? (see p. 5)
* Mini-lecture: Introduction to chi-square
* Challenge! Teams are presented with data we used earlier in the semester for an activity on distributions and hypothesis testing. The data are most appropriately analyzed using chi-square. I give the teams a few minutes to see if they can figure out what kind of chi-square to use, and return to the problem later.
* Mini-lecture: Chi-square goodness of fit test
* Students work on a goodness of fit test problem (done by hand) within their teams; we discuss the results as a class.
* Mini-lecture: Chi-square test of independence
* Students work on the challenge problem (done by hand) within their teams; we discuss the results as a class.
* Mini-lecture: Assumptions of chi-square
* Mini-lecture: Effect size measures
* Students work on effect size problems (done by hand) within their teams; we discuss the results as a class.
* Students work on two problems using the computer and SPSS
* Final team activity: “Meditation increases compassionate responses to suffering” (see pp. 7-8)

**Readiness Assurance Test**

1. In a chi-square test, the variables are:

a. interval level.

b. rank-order (ordinal).

c. ratio scale.

d. categorical (nominal).

2. The values in a chi-square distribution are always greater than 0 and:

a. less than 1.

b. are normally distributed.

c. can be quite large.

d. are negatively skewed.

3. The main idea of a chi-square test is that you:

a. test the estimated degree of fit (proportion of variance accounted for) of one variable to the other variable.

b. test how well the pattern of observed frequencies fits some expected pattern of frequencies.

c. compare the estimated population variances, to see if they vary from each other more than by chance.

d. compare the estimated population means, to see if they vary from each other more than by chance.

4. The degrees of freedom for the chi-square goodness of fit test are the:

a. mean number of individuals per category, minus 1.

b. number of categories minus 1.

c. mean number of individuals per category, minus the number of categories.

d. total number of individuals, minus the number of categories.

5. In a chi-square goodness of fit test, the null hypothesis is that the:

a. number of people in one category is no greater than the number of people in the other categories.

b. variances of the populations of categories are the same.

c. means of the populations of categories are the same.

d. observed proportion of people over categories does not depart from what is expected by chance.

6. In chi-square, the expected frequencies are the frequencies we would expect if the:

 a. null hypothesis is true.

b. null hypothesis is false.

c. research hypothesis is true.

d. research hypothesis is false.

7. Contingency table analysis is sometimes called the chi-square test of independence. This is because, under the null hypothesis, the:

a. variables are not related.

b. variables in the table should never be compared.

c. observed frequencies are larger than the expected frequencies.

d. distribution of one variable varies over different levels of the other.

8. The degrees of freedom for a contingency table analysis is the:

 a. total number of category levels minus 1.

b. number of categories minus 1.

c. number of participants minus the number of cells.

d. number of columns minus 1, times the number of rows minus 1.

9. One advantage of the chi-square test over most other inferential statistical procedures is that it:

a. can use the distributions of any other statistical procedure as a comparison.

b. has minimal assumptions about populations.

c. can be easily applied to repeated-measures designs.

d. does not require as many participants.

10. Which of the following would ***NOT*** be allowed in an ordinary application of the chi-square test?

a. The number of successes of an advertisement is compared for three different groups of people to see if the number of successes is equal for the three groups.

b. Tall and short people are compared on which religion they belong to.

c. The number of people of different ethnicities working in a particular company is compared to the proportions of those ethnicities in the general public.

d. Students' preference for studying while sitting or lying down is assessed as sophomores and then again as seniors.

**Introductory application activity**

For this activity, each team is given a sheet with the following scenarios and response options, as well as a small whiteboard and marker. Teams are given a couple of minutes to read each scenario and there is a race to see which team will hold up its whiteboard with an answer first. When there is disagreement, teams are asked to defend their answer. This activity generates a great deal of energy and gets students ready to explore topics in greater depth.

**What kind of chi-square?**

1. A gym wanted to assess member satisfaction with a new fitness program. They compared members who had joined the fitness program to members who had not joined the fitness program by asking them whether they were satisfied or not satisfied with their weight loss in the last six months. They hypothesized that members who had joined the fitness program were more likely to be satisfied with their weight loss.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

2. A pharmaceutical company wants to determine whether a sleeping pill is effective. They randomly assign individuals either to take a sleeping pill or to take a placebo. They compare the amount of time participants are asleep. They hypothesize the sleeping pill group will sleep longer than the placebo group.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

3. In a poll 200 residents in a small town were asked whether they supported the construction of a new mall. They responded whether they did, they didn’t, or whether they had no preference.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

4. A company that stocks a vending machine wants to optimize sales. The company offers five different beverage options and wants to know whether some beverages are more popular than others. To determine this, the numbers of beverages of each kind sold in one week are counted.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

5. An ad agency wants to see if a set of complimentary pens they’ve designed are gender neutral. They place both their green and yellow pens in a can and keep track of how many males and females pick up each color of pen.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

6. A researcher is testing how effective a list of words is at causing subjects to recall a specific word that was not listed. The researcher collects data from males and females during their freshman year, and then again during their sophomore year and then compares each person's data from the two years.

a. Chi-square goodness of fit test

b. Contingency table analysis (chi-square test of independence)

c. This can’t be analyzed using a chi-square!!

**Final application activity**

For this activity, students are given the introduction and method section of the following article (see file meditation\_compassion) : Condon, P., Desbordes, G., Miller, W. B., & DeSteno, D. (2013). Meditation increases compassionate responses to suffering. *Psychological Science, 20,* 1-3.

After reading these short sections, students complete a worksheet as a team, which I collected at the end of class. This article works perfectly for two reasons. First, it only reports and discusses two statistics, both of which appear in this module (chi-square and odds ratios). Second, after taking a contemplative pedagogy workshop in August, I integrated mindfulness techniques through the entire semester of this particular class, so students were primed to be interested in the topic.

While this is not a traditional 4S activity, it pulled several pieces of the module and course, in general, together and was a good ending to the topic.



**Team:**

**Meditation increases compassionate responses to suffering**

1. Read the introduction and method sections of the article.

2. What is the research hypothesis (is there more than one hypothesis)?

3. What is the null hypothesis?

4. The authors did not find a difference in behavior as a function of meditation protocol; therefore, the analysis uses data that collapsed participants in the two meditation conditions. The results of the study are below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Meditation training | Waiting-list control | Total |
| No help | 10  | 16  | 26 |
| Help | 10  |  3 | 13 |
| Total | 20 | 19 | 39 |

a. What type of analysis should be done on these data? Perform the analysis; report and interpret your result below.

b. Calculate an *r* measure of effect size; report and interpret the result.

c. Calculate the odds ratio (a *d* measure of effect size); report and interpret the result.

5. What are you overall conclusions about this study?

6. Do you notice anything interesting about the data? What questions are you left with?